Regulatory Toxicology and Pharmacology 81 (2016) 512–521

Contents lists available at ScienceDirect

Regulatory Toxicology and Pharmacology

journal homepage: www.elsevier.com/locate/yrtph

The need for transparency and reproducibility in documenting values for regulatory decision making and evaluating causality: The example of formaldehyde

Cynthia Van Landingham [a], Kenneth A. Mundt [b], Bruce C. Allen [c], P. Robinan Gentry [a, *]

[a] Ramboll Environ US Corporation, 3001 Armand St., Suite I, Monroe, LA 71201, United States
[b] Ramboll Environ US Corporation, 28 Amity St., Suite 2A, Amherst, MA 01002, United States
[c] Independent Consultant, 101 Corbin Hill Circle, Chapel Hill, NC 27514, United States

ARTICLE INFO

Article history:
Received 15 July 2016
Received in revised form
17 October 2016
Accepted 18 October 2016
Available online 19 October 2016

Keywords:
Reproducibility
Dose-response analysis
Formaldehyde
Inhalation unit risk estimates
Quantitative methods
Epidemiology

ABSTRACT

Reproducibility and transparency in scientific reporting is paramount to advancing science and providing the foundation required for sound regulation. Recent examples demonstrate that pivotal scientific findings cannot be replicated, due to poor documentation or methodological bias, sparking debate across scientific and regulatory communities. However, there is general agreement that improvements in communicating and documenting research and risk assessment methods are needed. In the case of formaldehyde, the peer-review conducted by a National Academy of Sciences (NAS) Committee questioned the approaches used by the Integrated Risk Information System (IRIS) in developing draft unit risk values. Using the original data from the key study (Beane Freeman et al., 2009) and documentation provided in the draft IRIS profile, we attempted to duplicate the reported inhalation unit risk values and address the NAS Committee's questions regarding application of the appropriate dose-response model. Overall, documentation of the methods lacked sufficient detail to allow for replication of the unit risk estimates, specifically for Hodgkin lymphoma and leukemias, the key systemic endpoints selected by IRIS. The lack of apparent exposure-response relationships for selected endpoints raises the question whether quantitative analyses are appropriate for these endpoints, and if so, how results are to be interpreted.

1. Introduction

Reproducibility and transparency in scientific research and reporting, both in the published literature and in documentation of decisions related to public health reached by authoritative bodies, have received significant discussion and debate (Bustin and Nolan, 2015; Campbell, 2014; Igbal et al., 2016; Jilka, 2016). The National Institutes of Health (NIH) are exploring ways to provide greater transparency of the data that are the basis for published manuscripts (Collins and Tabak, 2014) and have noted that the greater scientific community must take steps to correct this issue. In addition, recent commentaries and surveys highlight the growing lack of reproducibility in scientific research (Anonymous, 2016). One of the most immediate and impactful consequences for a lack of transparency or reproducibility is in the direct reliance on published but un-replicated scientific findings for human health risk assessment, including the derivation of cancer unit risk estimates.

In 2011, the National Research Council (NRC) of the National Academy of Sciences (NAS) convened a Committee to Review USEPA's Draft of the *Toxicological Review of Formaldehyde — Inhalation Assessment* in support of the Integrated Risk Information System (IRIS) (NRC, 2011). The Committee noted:

> *"Problems with clarity and transparency of the methods appear to be a repeating theme over the years, even though the documents appear to have grown considerably in length"*

A further review of the IRIS process in 2014 (NRC, 2014) noted progress in meeting the NRC (2011) recommendations, but further noted:

* Corresponding author.
E-mail address: rgentry@ramboll.com (P.R. Gentry).

http://dx.doi.org/10.1016/j.yrtph.2016.10.011
0273-2300/

"However, NRC committees have conducted several reviews of some of the more complex and challenging IRIS assessments in the last decade and have identified methodologic problems and pointed out deficiencies in EPA's approaches."

Formaldehyde provides one such complex database that introduces significant challenges for consideration in a standard IRIS assessment. It is an endogenously generated compound and, for selected endpoints, multiple studies provide inconsistent results, a few of which have suggested associations with formaldehyde exposure. Some have interpreted these findings (generally at face value and apart from the larger body of results) as reflecting causal associations. As an example, there has been much scientific debate regarding whether there is a causal association between formaldehyde exposure and selected lymphohematopoietic (LHP) endpoints, especially acute myeloid leukemia. Multiple authoritative bodies (IARC, 2012; NTP, 2014) have made hazard classification decisions (sufficient evidence in humans, known to be a human carcinogen) based on conclusions that the available evidence is sufficient to conclude that there is a causal association. For the LHP cancers, these conclusions have been based on the grouping of different types of cancers from a limited number of epidemiological studies (Zhang et al., 2009; Beane Freeman et al., 2009), with little or no consideration of findings reported in many other studies or the animal or mechanistic information, much of which lends no support for or even contradicts these conclusions. It is important to note that in reviewing the same critical studies for formaldehyde as IARC (2012) and NTP (2014), the European Chemicals Agency (ECHA, 2011) concluded that

"Altogether, in absence of convincing evidence for a biologically plausible mechanism and considering the discrepancy of results in epidemiological studies, a causal relationship between formaldehyde exposure and induction of myeloid leukaemia cannot be concluded."

The 2010 draft IRIS *Toxicological Review of Formaldehyde — Inhalation Assessment* provided the first quantitative estimates of a dose-response relationship between two lymphohematopoietic endpoints, Hodgkin lymphoma (HL) and all leukemias (combined category), and exposure to formaldehyde based on the results from a single epidemiological study (Beane Freeman et al., 2009). The use of these two endpoints by USEPA (2010) for the estimation of unit risk factors was based on the conclusion that the weight of the epidemiologic evidence supported a link between formaldehyde exposure and LHP cancers, particularly myeloid leukemias. In addition to HL largely being considered unrelated to environmental exposures, no other key epidemiological study demonstrates such an association, raising questions as to the validity of the finding in Beane Freeman et al. (2009). As for the combination of all leukemias, little scientific basis is provided for aggregating what increasingly are understood to be diverse diseases with different etiologies, prognoses and treatments.

In 2011, the NRC Committee review noted many uncertainties in the approach used by USEPA (2010) to estimate risk values. The Committee recognized that USEPA (2010) had relied upon selected associations reported between formaldehyde and various LHP cancers from a single study (Beane Freeman et al., 2009). The NRC (2011) Committee further recommended that USEPA conduct an independent analysis of the dose-response models to confirm the degree to which the models fit the data appropriately, as well as consider the use of alternative extrapolation models for the analysis of the cancer data. The NRC (2011) Committee concluded that this is especially important, given the use of a single study, the inconsistencies in the exposure measures, and the uncertainties associated with the selected cancers. In addition to the impact of these assumptions, the NRC (2011) Committee noted that while the National Cancer Institute (NCI) cohort studies, including Beane Freeman et al. (2009), may be the only studies with sufficient exposure and dose-response data needed for risk estimation, they are not without weaknesses and these need to be considered. This recommendation from the NRC (2011) Committee raised several challenges. While there is some guidance provided for the use of animal data for dose-response modelling (USEPA, 2012), the use of epidemiological data in the estimation of inhalation unit risk (IUR) estimates does not have guidance that provides a "road map" for conducting these types of assessments. When using epidemiological data for the estimation of unit risk values, more extensive documentation in the IRIS profile is needed to be able to clearly understand the data relied upon and the methods applied.

In a separate study (Checkoway et al., 2015), the raw data from the NCI cohort study (Beane Freeman et al., 2009) were obtained through a Technology Transfer Agreement (TTA) with the objective of replicating the findings reported by Beane Freeman et al. (2009), as well as conducting additional analyses not reported by Beane Freeman, specifically, acute myeloid leukemia (AML). The availability of these data provided an opportunity to attempt to replicate the unit risk estimates derived by USEPA (2010), as well as address some of the questions raised by NRC (2011). In addition, it offered the opportunity to conduct alternate independent analyses to evaluate specific leukemias, rather than all leukemias combined, and the impact of alternate dose-response models on the estimates of inhalation unit risk. The methods and results of the attempt to duplicate the USEPA (2010) unit risk values, as well as conduct alternate and independent analyses to address the questions raised by NRC (2011) are reported here.

2. Methods

2.1. Duplication of USEPA (2010) reported unit risks

Our goal was to follow the same process and methods used by USEPA (2010) in the estimation of unit risk factors for the two LHP cancers (Hodgkin Lymphoma and all leukemias (combined category)). However, as noted by NRC (2011), the documentation provided in USEPA (2010) related to the assumptions and processes used in the estimation of the unit risk values was limited. NRC (2011) has outlined five steps that it appears USEPA (2010) used in the estimation of formaldehyde unit risks:

1. Evaluate the association between formaldehyde exposure and LHP endpoints;
2. Convert the relative risk estimates into lifetime risk for the exposed population;
3. Compute lifetime risks for Hodgkin Lymphoma and/or all leukemia for the unexposed population;
4. Determine the maximum likelihood and lower bound estimates of the point of departure; and
5. Estimate inhalation unit risks.

Using these five steps, we attempted to duplicate the USEPA (2010) reported unit risks for Hodgkin lymphoma and "all leukemias" using the raw data from the Beane Freeman et al. (2009) study. In order to conduct this estimate, the following were needed:

- *An estimate of cumulative dose for each individual in the cohort.* This information was not provided in either USEPA (2010) or Beane Freeman et al. (2009) and must be determined from the raw data.

- *Person time at risk for each individual*. Also not provided in USEPA (2010) or Beane Freeman et al. (2009) and must be determined from the raw data.

Absent this necessary information and with no data available to confirm how it was used in estimating risk, assumptions were necessary that impact the estimation of parameters characterizing the relationship between dose and response.

NRC (2011) also recommended that the evaluation of the epidemiological data focus on the most specific diagnoses available. Based on this recommendation, analyses were conducted to include the consideration of individual LHPs rather than combination of endpoints (e.g. all leukemias) and evaluation of alternate dose-response models for these individual endpoints. While the impact of dose metric selection (e.g., 'peak'[1] versus cumulative) has been a point of discussion in interpretation of the NCI cohort (Checkoway et al., 2015), specifically the lack of actual peak measures or estimates, the USEPA (2010) has noted that cumulative exposure is generally the preferred metric for quantitative risk assessment and was relied upon for the estimation of unit risk values. Therefore, the analyses reported below focused on cumulative exposure estimates based on the data obtained through the TTA and reported in Beane Freeman et al. (2009) and Checkoway et al. (2015).

2.2. Evaluation of model selection

NRC (2011) noted that information was needed on the degree to which the model used (i.e., Poisson regression model) fits the data, especially for dose-response analysis. NRC (2011) further noted that this type of analysis is essential because dose-response models for risk estimation must fit the data well in the low-dose range and alternative extrapolation models, including Cox regression models and nonlinear model forms, should be considered in order to identify the best-fitting model. We conducted additional analyses to evaluate the potential impact of NRC (2011) comments on both the methods and the data relied upon for unit risk estimation, as well as consideration of multiple models. In addition to a Poisson regression model, the logistic regression model was considered, as well as a Cox regression model that was applied to the data from Beane Freeman et al. (2009) by Checkoway et al. (2015). All models used a 2-year lag for exposure, which is consistent with a lag considered by both Beane Freeman et al. (2009) and Checkoway et al. (2015).

A log-linear Poisson model, which is the model reported by Beane Freeman et al. (2009) to estimate the exposure-response relationship (β values), was used to compare the results in this analysis to the results published in Beane Freeman et al. (2009) in which the cumulative 2-year lag exposure variable was categorized into discrete exposure variables using the 4 categories reported (0 ppm-years, >0 and < 1.5 ppm-years, ≥1.5 and < 5.5 ppm-years, and ≥5.5 ppm-years). A log-linear Poisson model was also fit using the discrete dose categories reported by Checkoway et al. (2015) (<0.5 ppm-years, ≥0.5 and < 2.5 ppm-years, and ≥2.5 ppm-years). In addition, both a log-linear Poisson model and a logistic regression model were fit to the data using a categorization scheme for the 2-year lag cumulative dose that split the data into quartiles so that an approximately equal number of subjects were in each group (<0.05 ppm-years, ≥ 0.05 and < 0.4 ppm-years, ≥0.4 and < 2.4 ppm-years, and ≥2.4 ppm-years). All models were run considering person-time at risk, sex and race and adjusted for pay type (i.e., hourly vs. salary). For the logistic and Poisson models, quadratic terms for exposure were also considered. For evaluation of potential model fit to the data in the low concentration region, a visual examination of the Poisson and log-logistic model estimates were compared to the case status at the end of follow-up for each individual, again considering person-time at risk, sex, race and pay type.

[1] The 'peak' exposure metric used in Beane Freeman et al. (2009) is a relative peak estimator described in Stewart et al., 1986.

3. Results

3.1. Duplication of USEPA (2010) reported unit risks

3.1.1. Step 1 – evaluate the association between formaldehyde exposure and LHP endpoints

The attempt to estimate the unit risks reported in USEPA (2010) was initiated using the model parameters (β parameters from the log-linear Poisson regression model) provided to USEPA via personal communication by Dr. Laura Beane Freeman. The β parameters describe the relationship between exposure and response. Prior to estimating the unit risk, using the raw data, we attempted to replicate the model parameter estimates provided to the USEPA (2010) by Dr. Beane Freeman using log-linear Poisson regression, which is the same modelling approach reported to have been used to develop these estimates in both the Beane Freeman et al. (2009) publication and in the draft IRIS evaluation (USEPA, 2010) (Table 1). In addition, Cox and logistic regression models were considered.

Since cumulative exposure was the focus of the USEPA (2010) unit risk estimates, an initial analysis to evaluate the association between this exposure metric and the two endpoints relied upon for unit risk estimates (i.e., Hodgkin lymphoma and all leukemias combined) was conducted. Several variables were needed from the raw data, including the estimate of cumulative exposure (ppm) for each individual and person time at risk for each individual, neither of which are provided in USEPA (2010) or Beane Freeman et al. (2009) and had to be estimated from the raw data. In addition, in order to estimate the β parameters, the raw data regarding the number of deaths from a specific cancer and corresponding exposure metric must be divided into the same exposure quartiles as those reported by Beane Freeman et al. (2009) to evaluate the exposure-response relationship.

For the current analyses, the following steps were conducted to identify the data needed for analysis.

1. Using the work history data and date of birth, the data records were combined and organized to result in one or more record for each job so that no record spanned a calendar year or a change in age. Calculation of the duration of each work record was performed in this step with consideration of leap years. Since only start and stop months of work were provided in the raw data from Beane Freeman et al. (2009), the initial start and final stop day for a job were assumed to be the 15th of the month unless the start and stop months were the same month in the same year. In this case, the stop day was assumed to be the appropriate value for the end of the month (28, 29, 30 or 31). The gender, race, salary code and status of each individual (alive or dead) and cause of death ICD code were also attached to the individual's record.
2. The exposure and duration of exposure were summed over the months in a year when the individual was a specific age. During this step, the peak exposure category for each work record was determined.
3. The cumulative and lagged cumulative exposure and person-years of exposure were calculated.
4. The records were categorized into the strata of ranges of years (groups covering a 5 year period starting with 1960 and ending

Table 1
Comparison of modelling statistics from the current analysis to statistics reported in USEPA (2010).

	Current analysis												USEPA (2010)		
	Cox regression			Logistic regression					Poisson regression						
	p-value [a]	β (per ppm x year)	Standard error (per ppm x year)	R^2	LR p-value [b]	p-value [a]	β (per ppm × year)	Standard error (per ppm × year)	LR p-value [b]	p-value [a]	β (per ppm × year)	Standard error (per ppm × year)	p-value	β (per ppm × year)	Standard error (per ppm × year)
Hodgkin lymphoma (201)	0.013	0.0294	0.0119	0.0133	0.098	0.019	0.0288	0.0123	0.09	0.037	0.0243	0.0117		0.02959	0.01307
Leukemia (204–207)	0.058	0.0117	0.0062	0.0017	0.35	0.055	0.0121	0.00628	0.003	<0.001	0.0206	0.0057	0.08	0.01246	0.000642
Leukemia (204–207, excluding 204.1)	0.239	0.0092	0.0079	0.0011	0.64	0.206	0.01	0.00791	0.034	0.013	0.018	0.0073	–	–	–
Acute myeloid leukemia (205.0)	0.844	−0.004	0.0201	0.0016	0.82	0.869	−0.0032	0.0196	0.81	0.80	0.0045	0.0179	–	–	–

Cox regression model $h(t,x) = h_0(t)\ exp(\beta x + \gamma z)$.
Logistic regression model $Y = 1/[1 + exp(-a + \beta x + \gamma z)]$.
Poisson regression model $Ln(Y/t) = \alpha + \beta x + \gamma z$ OR $Y = t\ exp(\alpha) \times exp(\beta x)) \times exp(\gamma z)$.
Where Y is the expected number of events, α is the intercept, β is the slope term, x is the exposure, z is a covariate and t is the duration of exposure. In the Cox model h is the hazard rate.
[a] These p-values reflect the precision of any association between exposure and response, and show the probability that the beta value is not significantly different from zero. P-values < 0.5 indicate that the beta parameter is significantly different from zero.
[b] The likelihood ratio p-values of difference between a null and dose-dependent model (e.g. test of $\beta = 0$) where small p-values reject the hypothesis that $\beta = 0$.

with 2010), and age groups (groups covering a 5 year range starting with the age of 15 and ending with 85), where the lowest year group included all records prior to 1965, and the 1965 group included years 1965 through 1969, with all job records occurring in 2010 and after included in the 2010 category. For ages, all ages less than 20 were included with the 15 year old age group, and the second group labelled 20 included all ages from 20 through 29.

5. The final record for each individual included an indication of dead or alive. For those individuals who had died, the ICD codes were used to set up yes/no flags indicating whether Hodgkin lymphoma, leukemia or acute mylogenous leukemia were found in that individual.

This process resulted in 1,047,291 work records that were then used in the analyses. All analyses used stratification for age group, year group, gender and race, with all the models adjusted for salary type treated as a classification variable. The Poisson analysis (SAS Proc Genmod) used a Poisson distribution, a log link and an offset of the natural log of the cumulative person-years of exposure. SAS Proc Logistic was used to perform the logistic regression and Cox proportions hazards models were performed using STATA (Checkoway et al., 2015).

Beane Freeman et al. (2009) reported that the cut points for the exposure groups were based on the approximate 60th and 80th percentiles from the cumulative exposures for those subjects with cancer. In attempting to duplicate the number of cancers within each exposure group, the cut points of 1.5 and 5.5 ppm-years (cumulative exposure groups defined by Beane Freeman et al. (2009) as ≤0 to 1.5, 1.5 to <5.5, ≥5.5 ppm-years) could not be duplicated based on the estimated 60th and 80th percentiles using the raw data. The calculations for the current assessment resulted in the determination of 1.2 and 4.2 ppm-years as the 60^{th} and 80^{th} percentiles for the cumulative exposure of the subjects with cancer. In addition, the number of unexposed workers (4359) reported by Beane Freeman et al. (2009) could not be replicated. Using the raw data, only 2676 unexposed workers could be identified.[3]

Regardless of the lack of ability to duplicate this determination of exposure, an evaluation of the exposure-response relationship was conducted. For the "all leukemia" category, exposure-response was evaluated including and excluding chronic lymphocytic leukemia (CLL), because, as noted by Checkoway et al. (2015), CLL has been classified as a non-Hodgkin lymphoma (NHL) since 2001 (Muller-Hermelink et al., 2001; Campo et al., 2011).

Other models were attempted in this process. Using quadratic terms for exposure failed to provide any better fit of the models to the data. In addition, the effect of exposure to other substances were explored but these did not improve the model fits substantially, either.

As noted in Table 1, in attempting to duplicate the β parameter and standard error for each cancer type, similar values could be estimated, but the estimates reported in USEPA (2010) could not be duplicated, which can impact attempting to duplicate unit risk estimates. In addition, it is important to note that no significant association between leukemia as a class of diseases (p-values > 0.05; Table 1) or specifically for acute myeloid leukemia ($p \geq 0.8$) with cumulative exposure to formaldehyde was found (using the typical 0.05 as the determinant of "significant") for either the Cox regression or the logistic regression. In addition, the estimated β parameter for acute myeloid leukemia (~-0.004 from the Cox regression and the logistic regression) indicates that the slope is in the negative direction (decreasing incidence with increasing exposure). These results for AML suggest that it would not be appropriate to rely upon these negative data independently in the dose-response modelling for the estimation of a unit "protection" estimate. As imprecise positive estimates of a β parameter should not be interpreted as evidence of risk, imprecise negative β parameters should not be interpreted as beneficial or protective. For all the logistic models, the likelihood ratio test indicates that the β parameter is not statistically different from zero. Similarly the likelihood ratio test of the Poisson models for Hodgkin lymphoma and the acute myeloid leukemia also indicate that the β parameter is not statistically different from zero. Only for the Poisson models of combined leukemias are the β values considered to be statistically significantly different from zero. However, as these are

combined types of leukemia which are not recommended by the NRC (2011) and there is almost a factor of 2 difference between the β estimates between the different models in the current analysis and the USEPA (2010) β estimate, there is still large uncertainty in the results.

The estimated β parameter for Hodgkin lymphoma was comparable to that reported in USEPA (2010); however, there was a difference in the standard error and a larger difference in the p-values. USEPA (2010) reported a non-significant trend between cumulative formaldehyde exposure and Hodgkin lymphoma based on information reported in Beane Freeman et al. (2009), while the current analysis suggested a significant trend (p-value = 0.013). These results are consistent with those reported by Checkoway et al. (2015). However, Checkoway et al. (2015) notes that the increased risk of HL has not been observed in other occupational studies of formaldehyde-exposed cohorts, and is not regarded as plausibly related to environmental chemical exposures.

Because the β parameters could not be duplicated, it was concluded that while additional steps could be conducted to evaluate the transparency of the process, the lack of ability to duplicate this first step would result in a lack of ability to duplicate the reported unit risks. Even having access to the raw data from the Beane Freeman et al. (2009) study, there were not enough details regarding the methods used to evaluate the data provided in USEPA (2010) to duplicate the initial β parameters necessary to initiate the unit risk estimate process.

3.1.2. Step 2 – convert the relative risk estimates into lifetime risk for the exposed population

Relying strictly on the β parameters reported in USEPA (2010), even though they could not be duplicated, an attempt was made to conduct the remaining steps of the estimation of unit risk as outlined by NRC (2011). USEPA (2010) noted that the β parameters were used in a life table analysis to calculate lifetime extra cancer risks from formaldehyde exposure. This step, as well as step 3, requires the use of a life-table method in conjunction with (a) the Poisson model mortality risk, (b) age-specific all-cause mortality rate in the United States population, and (c) Hodgkin lymphoma and all leukemia mortality rates, all of which can be derived from the NCI's Surveillance, Epidemiology and End Results (SEER) database. SEER collects cancer incidence data from multiple geographical areas in the United States. This step also requires estimates of the effective concentration (EC) for occupational exposure adjusted to continuous ambient exposure (the standard exposure metric relied upon by USEPA in the estimation of a unit risk) by multiplying by the ratio of days in a year to work days (240, 50 weeks of 5 day work weeks) and the ratio of daily inhalation rate (20 m^3) to work day inhalation rate (10 m^3) (USEPA, 2010).

$$EC = exposure\ (ppm) \times \frac{365}{240} \times \frac{20}{10}$$

USEPA (2010) provided a spreadsheet (Appendix C of USEPA, 2010; Supplemental Tables S1 and S2) illustrating the life table used for the extra risk calculation for the derivation of the LEC_{0005} (95% lower confidence limit on the effective concentration corresponding to an extra risk of 0.05%) relied upon for estimating the IUR based on nasopharyngeal (NPC) mortality reported by Hauptmann et al. (2004). USEPA (2010) noted that the same general methodology described for NPC mortality estimates was used for Hodgkin lymphoma and leukemias, with the following exceptions:

- U.S. age-specific 2006 all-cause mortality rates (NCHS, 2009);
- NCHS age-specific 2002–2006 background mortality rates for Hodgkin lymphoma and leukemia (http://seer.cancer.gov/csr/1975-2006/) for all race and gender groups; and
- A 2-year lag period instead of a 15-year lag period.

It is important to note that USEPA (2010) provided no citation for the NCHS (2009) all-cause mortality rates, so it was assumed this was obtained from the NCHS website (http://www.cdc.gov/nchs/data/nvsr/nvsr57/nvsr57_14.pdf) as the background mortality rates for specific cancers (Heron et al., 2006). While this does provide data needed to allow the assessor to attempt to duplicate this procedure, there is no comparable life-table for Hodgkin lymphoma or all leukemias to ensure that comparable results are achieved. Relying upon these sources and following these approaches, the IURs provided in USEPA (2010) could not be duplicated using the reported sources and methodology. This was also true for NPC for which the life table was provided (Appendix C; USEPA (2010)). In attempting to duplicate the IURs reported for NPC, it was determined that the values reported from the use of the life table instructions provided could not produce the reported IURs for NPC (see supplemental Table S1 for the re-creation of the calculations that would correspond to the unit risks reported in USEPA (2010) when using the instructions provided by USEPA (2010) for Table C-1. The difficulty in duplicating the life table reported was related to the function reported for estimating the NPC incidence hazard rate (Column L in Supplemental Table 2). Using the USEPA (2010) β of 0.0518 (SE 0.01915) and the calculations as specified in Table C-1 of USEPA (2010), the estimated EC_{0005} and LEC_{0005} would be 0.103 and 0.0623 ppm, respectively, with a unit risk of 8×10^{-3}. However, the calculations specified in Appendix C of USEPA (2010) indicated a function for the hazard incidence rate of $hx_i = h_i \times (1 + \beta \times xdose)$ which is inconsistent with the model of risk that was used to determine the β value ($RR = e^{\beta X}$, where β represents the regression coefficient for exposure and X is exposure as a continuous variable) (USEPA, 2010). When the hazard rate function is changed to $hx_i = h_i \times (e^{\beta \times xdose})$ to properly reflect the underlying risk function, the values estimated by the revised life table were the same as those reported by the USEPA in Tables 5–11 for EC_{0005} and LEC_{0005} based on NPC incidence for formaldehyde exposure (0.074 and 0.046 ppm, respectively, see supplemental Table S3 for the adjusted life-table calculation). However, it is important to note that these estimates rely upon the β parameters reported in USEPA (2010), which cannot be duplicated.

3.1.3. Step 3 – compute lifetime risks for Hodgkin Lymphoma and/or all leukemia for the unexposed population

As noted in USEPA (2010), USEPA cancer risk estimates are typically derived to represent a plausible upper bound on increased risk of cancer *incidence*, typically based on experimental animal incidence data. However, epidemiological studies more often present results based on mortality data, which is true for the Beane Freeman et al. (2009) study. For cancers with low survival rates, mortality-based estimates are a reasonable approximation of cancer incidence risk. However, USEPA (2010) largely documents its approach to the evaluation of nasopharyngeal cancers and noted the need to estimate incidence-based risks. Estimation of the incidence of a particular cancer type using mortality data can be conducted by acquiring the age-specific incidence rates for a specific cancer from the SEER program. In order to estimate the potential risk of incidence of a cancer type, the data from the SEER database are used to adjust the mortality data assuming that the exposure-response relationship for incidence and mortality of a cancer type are the same. An examination of the assumptions and adjustments made to the Beane Freeman et al. (2009) data for lymphohematopoietic cancers follows.

Table 2
Extra risk estimates for Hodgkin lymphoma mortality from various levels of continuous exposure to formaldehyde (reproduced from Tables 5–14 in USEPA (2010)).

Exposure concentration (ppm)	As reported by USEPA (2010)		Estimated using the life table provided in USEPA (2010) [a] with adjustments to the hazard function	
	Extra risk	95% UCL on extra risk	Extra risk	95% UCL on extra risk
0.0001	2.04×10^{-7}	3.53×10^{-7}	2.52×10^{-7}	4.36×10^{-7}
0.001	2.05×10^{-6}	3.55×10^{-6}	2.53×10^{-6}	4.38×10^{-6}
0.01	2.10×10^{-5}	3.71×10^{-5}	2.59×10^{-5}	4.59×10^{-5}
0.1	2.79×10^{-4}	6.17×10^{-4}	3.44×10^{-4}	7.63×10^{-4}
1	1.63×10^{-1}	8.36×10^{-1}	1.90×10^{-1}	8.53×10^{-1}
10	9.89×10^{-1}	9.90×10^{-1}	9.89×10^{-1}	9.90×10^{-1}

[a] Using the supplied information in the life table provided in USEPA (2010) with an adjustment in column L for the incidence hazard rate in interval l ($hxi = hi \times e^{(\beta \times xdose)}$) for the estimates of $\beta = 0.02959$, SE = 0.01307.

Table 3
Extra risk estimates for leukemia mortality from various levels of continuous exposure to formaldehyde (reproduced from Tables 5–15 in USEPA (2010)).

Exposure concentration (ppm)	Calculated by USEPA (2010)		Estimated using the life table provided in USEPA (2010) [a] with adjustments to the hazard function	
	Extra risk	95% UCL on extra risk	Extra risk	95% UCL on extra risk
0.0001	1.64×10^{-6}	3.02×10^{-6}	1.65×10^{-6}	3.06×10^{-6}
0.001	1.64×10^{-5}	3.03×10^{-5}	1.65×10^{-5}	3.07×10^{-5}
0.01	1.66×10^{-4}	3.10×10^{-4}	1.67×10^{-4}	3.13×10^{-4}
0.1	1.87×10^{-3}	3.90×10^{-3}	1.89×10^{-3}	3.95×10^{-3}
1	8.07×10^{-2}	5.19×10^{-1}	8.16×10^{-2}	5.28×10^{-1}
10	9.80×10^{-1}	9.89×10^{-1}	9.80×10^{-1}	9.89×10^{-1}

[a] Using US 2006 mortality rates, the adjusted life table structure and potency estimates ($\beta = 0.01246$, SE = 0.006421) from USEPA (2010).

Table 4
Relative risk based on peak exposure from Poisson model stratified by calendar year, age, sex and race and adjusted for pay category.

	0 ppm		>0 to <2.0 ppm		> = 2.0 to <4.0 ppm		> = 4.0 ppm		Log likelihood	p-value
Total in group	3139		10,302		6010		6168			
Person-years	104,386		415,987		254,723		256,618			
	Cases	RR (95% CI)	Cases	RR (referent)	Cases	RR (95% CI)	Cases	RR (95% CI)		
Hodgkin lymphoma (201)	2	3.32 (0.60–18.26)	6	1.0	8	0.76 (0.30–1.89)	11	2.96 (0.94–9.27)	−309.87	0.04
Leukemia (204–207)	7	1.83 (0.76–4.40)	41	1.0	27	0.58 (0.36–0.93)	48	0.58 (0.36–0.93)	−1177.94	0.004
Leukemia (204–207, excluding 204.1)	6	1.61 (0.61–4.24)	28	1.0	20	0.56 (0.32–0.96)	37	1.17 (0.65–2.09)	−901.65	0.009
Acute myeloid leukemia (205.0)	4	1.21 (0.33–4.43)	9	1.0	9	0.77 (0.32–1.84)	12	1.72 (0.67–4.43)	−374.47	0.34

Since USEPA's life table analysis relied upon background mortality rates to determine the extra risk from the incidence of the endpoint of interest, the effect of using background incidence data for Hodgkin lymphoma and all leukemia was explored. The background mortality rates were adjusted to reflect the background incidence of the endpoint by replacing the mortality rate attributed

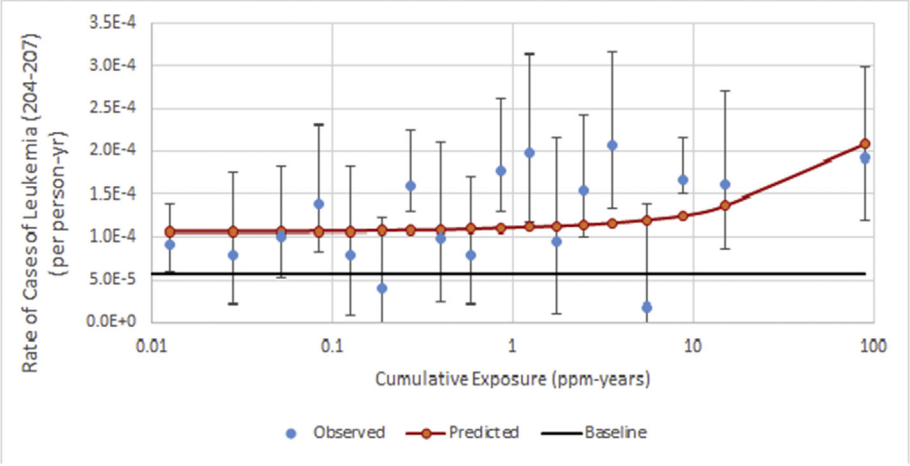

Fig. 1. Comparison of estimated cases from the Poisson regression model to number of cases of leukemia observed at the end of follow-up period in the Beane Freeman et al. (2009) study. Observed and predicted results over full observed exposure range.

to that endpoint with the incidence rate of that endpoint. Making this correction resulted in a difference of between 10 and 21% in the estimated risks for the current analysis.

3.1.4. Step 4 – determine maximum likelihood and lower bound estimates of point of departure

USEPA's carcinogenicity risk-assessment guidelines (USEPA, 2005) recommend the use of an extra risk of 1–10% for deriving effective concentration at the Point of Departure (POD), or for the USEPA (2010) IRIS assessment. NRC (2011) noted that in USEPA (2010) there was an unusual choice of a 0.05% extra risk for Hodgkin lymphoma and 0.5% extra risk for all leukemias. USEPA (2010) noted the issues with using standard extra risk levels (e.g., 10%) in that the risks using these standard extra risk assumptions resulted in relative risk estimates that were substantially higher than those observed in the epidemiology study. Therefore, the choice of the extra risk value to use was based on the background mortality rate for each individual cancer type compared to the relative risk estimates observed in the Beane Freeman et al. (2009) study. Relative risk estimates were determined starting at the 10% extra risk level, decreasing the extra risk level until the relative risk estimates were within the observable range of the epidemiology study. For example, if the 1% level of risk associated with the relative risk estimates for NPC were higher than those observed in the Beane Freeman et al. (2009) study, the extra risk level of concern was lowered until the relative risk estimates were below the relative risk estimates from the Beane Freeman et al. (2009), so an upward extrapolation could be conducted. This approach effectively assumes that nothing observed in the Beane Freeman et al. (2009) could be attributable to background incidence of these cancer types.

Using the hazard rate function as instructed in the life table example (Footnote for Column L, Table C-1 of USEPA (2010)), the extra risk and 95% upper confidence limits on extra risk provided in USEPA (2010) cannot be reproduced (Tables 2 and 3). However, using a life table that had a hazard rate function consistent with the underlying risk function produced results that were similar to those reported by the USEPA (2010). Supplemental Tables S2 and S4 show the differences in the risk values calculated at an exposure of 1 ppm using the USEPA (2010) instructions (Table S2) versus the revised life table (Table S4) with the modified hazard function that was necessary to duplicate the EC, LEC and unit risk values reported in USEPA (2010). While there was some correspondence, there were still some differences in the values that were calculated for the extra risk (Tables 2 and 3) and there is some concern about the appropriateness of the risk estimates, especially large estimates of risk for values above 0.1 ppm. An exposure of 0.1 ppm is within the range of exposures (0.01–4.3 ppm – TWA) reported by Beane Freeman et al. (2009). The relative risk values estimated for these exposures approach 100% and are inconsistent with the observed incidences of cancers in the Beane Freeman et al. (2009) study.

3.1.5. Step 5 – convert the relative risk estimates into lifetime risk for the exposed population

With the results from step 4, the lower bounds on exposure (LECx) and the extra risk level should then be used to determine the unit risks. However, because the model parameters from step 1 could not be replicated, an attempt was made to replicate the MLE and lower bounds using the USEPA (2010) reported model parameters. Using a life table analysis that follows the methods provided in Appendix C of USEPA (2010) and the reported model parameters, the MLE and lower bounds on dose for Hodgkin lymphoma and all leukemia could not be replicated. Using the available parameters and results reported in USEPA (2010) and using the USEPA's parameters, a 12–27% difference in unit risk values was determined for leukemia, Hodgkin's lymphoma and NPC from those reported by the USEPA (2010). However, when the life table was adjusted to be consistent with the relative risk model that was the basis of the β value used in USEPA (2010), the values reported by the USEPA could be replicated.

In noting the potential differences in unit risk estimation, this 12–27% difference could be considered in combination with the potential differences in unit risk from step 1 (differences in the model results), as well as the potential impact of the differences in risk from step 3. Therefore, the inability to replicate individual steps in the process may result in unit risk estimates different from those in USEPA (2010) by 100% or greater due to differences in the slope factors (up to 100% difference) as well as differences in life table analysis results (12–27%) that would be calculated following the documentation provided in USEPA (2010).

Analyses were also conducted using the "peak" exposure metric, rather than the continuous metric relied upon by USEPA (2010) for their evaluation. This was conducted using the same model (log-linear Poisson stratified by calendar year, age sex, and race and adjusted for pay category) as Beane Freeman et al. (2009), but in contrast to the results reported by Beane Freeman et al. (2009), no significant relative risks were estimated (Table 4). Reasons for the differences between the current analyses and those reported by Beane Freeman et al. (2009) could include that the specific dates of job start and job end were not provided, nor were the specific dates that follow-up started or ended: only month and year were reported.

3.1.6. Evaluation of model selection

In evaluating the potential fit of the model to the data, there are various tests that can be performed to look at the predictive power of a model (e.g. R^2 tests, χ^2 tests), to make comparison between models (e.g. AIC and other log-likelihood tests) or graphical representations of the data to visualize the fit. However, since no such statistics were provided in either Beane Freeman et al. (2009) or USEPA (2010), comparisons can only be made among the models fit to the data in this current analysis. The R^2 values reported for the logistic regression performed in this analysis were uniformly poor (i.e., 0.05 or less) indicating poor predictive ability of the models. For the Poisson models, there were small values for the Pearson χ^2 value which, with the large sample size, achieved a better fit to the data (p-values close to 1). However, in graphs presented in this analysis using the data at the end of follow-up, the rate of all leukemias was plotted against the continuous exposure as well as the model predicted rates estimated for both the Poisson regression model (Fig. 1) and the logistic model (Fig. 2).[2] These figures show large variability in the observed rates in the low concentration region which subsequently makes comparison and evaluation of the fit of the model to the data difficult. This variability also makes any predictions made with models fit to these data highly uncertain. In addition, the predictions of extra risk provided by USEPA (2010)

[2] The graphs were constructed using the 5% percentiles (e.g. 5%, 10%, 15%, etc.) of the cumulative exposure, and sums of the person-years, number of individuals and number of observed and predicted leukemias per percentile to determine the rates. The confidence limits for the logistic graph were calculated using binomial confidence limits on the observed rates of leukemia per percentile group of exposure, and the Poisson confidence limits are exact confidence limits based on the Poisson distribution.

[3] This number of unexposed workers identified in the current analysis (2676) is consistent with the number determined by Checkoway et al. (2015) in a separate reanalysis of the raw data from Beane Freeman et al. (2009) study. When this difference was discovered by Checkoway et al. (2015), communications with Dr. Beane Freeman indicated that the number of unexposed workers reported was a mistake and should have been 3,108. However, Checkoway et al. (2015) could not duplicate this number of unexposed workers either using the raw data.

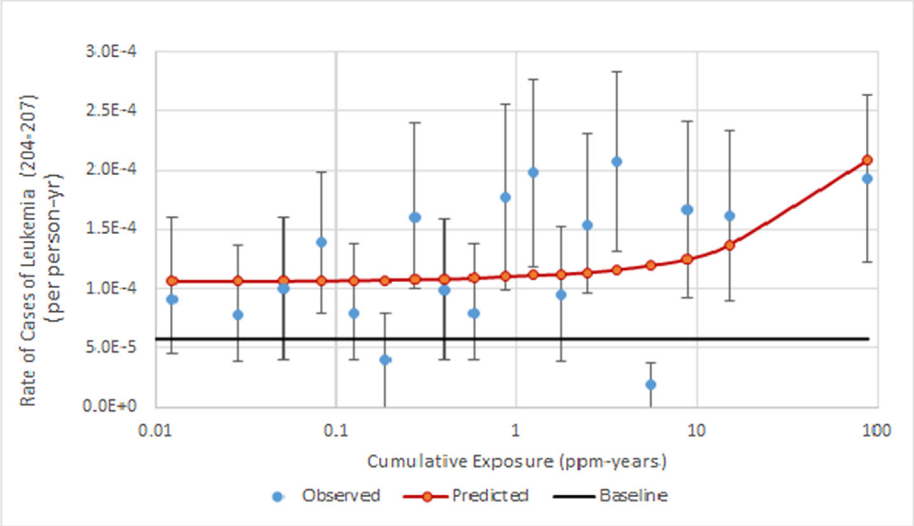

Fig. 2. Comparison of estimated cases from the logistic regression model to number of cases of leukemia observed at the end of follow-up period in the Beane Freeman et al. (2009) study. Observed and predicted results over full observed exposure range.

associated with higher concentrations (1 and 10 ppm) are above the observable range and involve upward extrapolation. The results are estimates of extra risk approaching 1, which are unreasonable.

While each model provides predictions that "run through the middle" of the data, it is clear that neither model can adequately predict the exposure-response relationships or lack of pattern in the lower concentration region (Figs. 1 and 2), as the data in this region of the exposure-response curve appears to be comparable to random variation. In the low concentration region, the data lack a clear monotonic dose-response relationship, which may explain lack of a significant trend ($p = 0.08$) even for the combination of all leukemias. Overall, the models do not fit the pattern of exposure-response in the data. While the models appear to be more consistent with the data at concentrations greater than 10 ppm-years, this comparison is largely influenced by two data points. It is possible that this shape of the exposure-response curve may explain the unusual nonlinearities in the estimates of extra risk provided by USEPA (2010) (Tables 2 and 3). However, explaining this unusual exposure-response behavior is difficult due to the inability to duplicate the unit risk estimates provided in USEPA (2010).

4. Discussion

One of the greatest challenges in attempting to duplicate unit risk factors estimated by USEPA is attempting to duplicate those specifically based on epidemiological data. When USEPA has relied upon animal data for the estimation of unit risk values, even when the documentation provided is limited, there are guidelines available (USEPA, 2012) that provide specific steps and assumptions used by USEPA in the dose-response analysis of animal data. However, when epidemiological data are applied, there is not comparable guidance, and the necessary additional detail may not be provided in the IRIS documentation to allow for transparency and the ability to duplicate risk values.

In the case of formaldehyde, the draft IRIS toxicological review (USEPA, 2010) provided documentation largely on the estimation of IURs from the cases of NPC from the NCI cohort reported by Hauptmann et al. (2004), assuming that these methods could easily be extended in an attempt to duplicate values for lymphohematopoietic cancers provided in an update to the NCI cohort by Beane Freeman et al. (2009). The results from this assessment, in attempting to duplicate unit risk values for lymphohematopoietic cancers, demonstrate that this is not the case.

Difficulty in duplication of results from each step of the process of the estimates of IURs, following the steps as outlined by NRC (2011), started with the initial step that involved duplication of the β parameters from the log-linear Poisson regression model as provided by Dr. Laura Beane Freeman to the USEPA. In the initial step of the process, our results suggest no significant association between cumulative exposure to formaldehyde, which is the exposure metric relied upon by USEPA (2010) for the estimation of the IURs, and either all leukemias combined or acute myeloid leukemia specifically. This lack of association is directly relevant to evaluation of causality and should be considered earlier in the determination of what endpoints likely are caused by exposure to formaldehyde and therefore which associations might be relied upon for the estimation of IURs. Based on the results for all leukemias, as well as AML, with no significant trends observed, it is not appropriate to conduct dose-response modelling only on null findings. In addition, while similar, the β values could not be duplicated even with the availability of the raw data, which suggests that the methods applied are not adequately documented in USEPA (2010).

USEPA (2010) relied heavily upon the Beane Freeman et al. (2009) study for risk estimation associated with lymphohematopoietic tumors, with the NRC (2011) committee noting that this may be the only study with sufficient exposure and dose-response data needed for risk estimation. However, they also noted that this study is not without weaknesses and these need to be considered. A reanalysis of the raw data from the NCI study (Beane Freeman et al., 2009) was conducted by Checkoway et al. (2015). While basic results were replicated, additional analyses of the associations of specific lymphohematopoietic cancers, specifically acute myeloid leukemia (AML) with various metrics of formaldehyde exposure (peak, average, cumulative) and using a more standard definition of peak exposure than that relied on by Beane Freeman et al. (2009) were reported. The re-evaluation highlighted many of the limitations in the data from this cohort, and the new analyses indicated no clear association with AML. It is not clear why AML results had not been reported in any of the updates of this study, and not considered in the IRIS evaluation, given that AML has been highlighted as the lymphohematopoietic cancer most likely to be

relevant to a chemical agent, primarily based on its association with benzene.

The results from the current analysis for Hodgkin lymphoma also provide estimates inconsistent with those reported by USEPA (2010). Using the cumulative exposure metric, USEPA (2010) reported no significant trend for Hodgkin lymphoma. The current analysis suggests a significant trend (Table 1 – p = 0.013), which is consistent with the results from Checkoway et al. (2015) reporting increased relative risk estimates for Hodgkin lymphoma in the highest exposure categories of cumulative and peak exposures. As noted in Checkoway et al. (2015), these findings are complicated because there is little epidemiological support for chemical exposures in the etiology of Hodgkin's lymphoma. There is an absence of an increased risk for this cancer type in other occupational cohorts, as well as the lack of a plausible biological mechanism. In addition, NTP (2014) noted that because the evidence for Hodgkin lymphoma is mainly limited to the NCI cohort study, a causal association is not established. As with all leukemias, including AML, there are questions related to a causal association between cumulative formaldehyde exposure and this cancer type that suggest that the estimation of a quantitative measure of risk using these data are inappropriate.

NRC (2011) also highlighted that the modes of action for formaldehyde-induced Hodgkin lymphoma and for leukemias have not been established. Moreover, the studies that demonstrate the lack of systemic delivery of formaldehyde following inhalation exposure (Lu et al., 2011; Moeller et al., 2011; Edrissi et al., 2013; Yu et al., 2015) draw into question the biological plausibility of formaldehyde causing any LHP cancer. NRC (2011) noted that

> *"Although EPA postulated that formaldehyde could reach the bone marrow either as methanediol or as a byproduct of nonenzymatic reactions with glutathione, numerous studies described above have demonstrated that systemic delivery of formaldehyde is highly unlikely at concentrations below those which overwhelm metabolism according to sensitive and selective analytic methods that can differentiate endogenous from exogenous exposures."*

Thus, substantial uncertainties remain in using both Hodgkin lymphoma and leukemias (all or individual) for consensus cancer risk estimation. Formaldehyde is rapidly metabolized and highly reactive and, because it is an endogenous compound, a detectable change in the natural background or endogenous levels would need to occur in order to result in the potential for adverse effects. Multiple studies using multiple species, including non-human primates, have been conducted using a sensitive analytical method that can measure endogenous versus exogenous formaldehyde DNA adducts (Yu et al., 2015; Edrissi et al., 2013; Moeller et al., 2011; Lu et al., 2011). The results of these studies indicated that inhaled formaldehyde was found to reach nasal respiratory epithelium, but not other tissues distant to the site of initial contact. These results suggest a lack of an ability for exogenous or inhaled formaldehyde exposure to affect endogenously present concentrations of formaldehyde.

Although the Draft Review cites hypotheses proposed by Zhang et al. (2010) regarding the theoretical development of leukemia following inhalation of formaldehyde, there is no documented evidence to support the validity of these hypotheses. In fact, Zhang et al. (2010) note that their hypotheses related to mechanisms of leukemia clearly require additional testing. The existing mechanistic data for formaldehyde provide no evidence that exogenous formaldehyde will be transported from the point of contact to distant sites, but do provide evidence that formaldehyde does not affect the relevant target cells for leukemia (bone marrow or peripheral blood) (Yu et al., 2015; Edrissi et al., 2013; Moeller et al., 2011; Lu et al., 2011).

Overall, the documentation of the methods applied by USEPA lacks sufficient transparency and detail for duplication of the unit risk estimates provided in USEPA (2010), even with the availability of the raw data from the Beane Freeman et al. (2009) study that USEPA relied upon for estimation of the risk of Hodgkin lymphoma or all leukemias. This lack of transparency and detail may result in different estimates of unit risks, including invalid estimates, especially as initial analyses resulted in a lack of a significant dose-response relationship for selected endpoints.

In attempting to duplicate the USEPA (2010) calculations, difficulties were encountered at each step, largely due to a lack of critical information provided in the IRIS documentation. Even though analyses were conducted multiple times with different assumptions, all of which could be consistent with the description provided by USEPA (2010), the unit risk values could not be duplicated. The results of the analyses yielded conflicting and different estimates with each step of the analysis, with differences in each step up to a factor of 2. The inability to replicate individual steps in the process may result in unit risk estimates different from those in USEPA (2010) by 100% or greater due to differences in the slope factors (up to 100% difference) as well as differences in life table analysis results (12–27%). Perhaps most problematic, the first step of the analysis did not determine significant exposure-response relationships between formaldehyde and LHP endpoints for the metric (cumulative exposure) needed in the estimation of an IUR. The resulting analysis, while it can be mechanically performed, provides no valid or useful insights on the risks of formaldehyde exposure. Regulatory dependence on these analyses may therefore lead to erroneous guidance, policies and laws.

These results highlight the necessity of clear and transparent reporting of both methods and data used in the estimation of unit risk values. Values provided by the IRIS program of USEPA are relied upon by other federal and state agencies in regulatory decision-making related to the development of standards and guidelines for environmental, consumer product and workplace exposure to chemicals. The inability to duplicate these types of values only escalates the scientific debate over the applicability of these standards and the scientific data necessary to support conclusions regarding acceptable levels of human exposure to chemicals.

Acknowledgements

This project was supported by funding provided by the Research Foundation for Health and Environmental Effects (Contract number R0072). RFHEE is a 501(c) (3) tax-exempt organization established by the American Chemistry Council (ACC). Established in 1995, RFHEE works to address uncertainties and answer questions on health and environmental issues related to the production and use of chemistry. RFHEE works with industry, public agencies, academia and other foundations on projects that will furnish crucial information, peer-reviewed scientific research, expert panels, and workshops that will provide information for public policy. RFHEE receives contributions from government, industry (companies and trade organizations), and other industry partners who share its commitment to improving the factual and scientific understanding about critical questions associated with and/or impacting chemistry. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of any manuscript. The funders were given the opportunity to review draft manuscripts at the time of external peer review. The purpose of such review was to allow input on the clarity of the science presented but not on interpretation of the research findings. The researchers' scientific conclusions and professional judgments were not subject to the funders' control.

Most of the authors (Gentry, Van Landingham, and Mundt) are employees of Ramboll Environ US Corporation, and performed this work as part of their normal employment. Ramboll Environ US Corporation is a consulting firm providing services in environmental and health sciences matters to private firms, trade organizations, and government agencies. Mr. Allen provided senior technical review on the manuscript and its content as an independent consultant with fee for service. The authors had sole responsibility for the analyses performed, the interpretations made, conclusions drawn and the writing of the paper, which may not necessarily reflect the views of RFHEE. None of the authors have appeared as experts in any litigation related to formaldehyde.

Appendix A.. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.yrtph.2016.10.011.

Transparency document

Transparency document related to this article can be found online at http://dx.doi.org/10.1016/j.yrtph.2016.10.011.

References

Anonymous, 2016. Reality check on reproducibility. Nature 533 (7604), 437 [Editorial].

Beane Freeman, L.E., Blair, A., Lubin, J.H., Stewart, P.A., Hayes, R.B., Hoover, R.N., Hauptmann, M., 2009. Mortality from lymphohematopoietic malignancies among workers in formaldehyde industries: the National Cancer Institute Cohort. J. Natl. Cancer Inst. 101 (10), 751–761.

Bustin, S.A., Nolan, T., 2015. Improving the reliability of peer-reviewed publications: we are all in it together. Biomol. Detect. Quantification 7, A1–A5.

Campbell, H., 2014. The Corruption of Peer Review Is Harming Scientific Credibility. Dubious Studies on the Danger of Hurricane Names May Be Laughable. But Bad Science Can Cause Bad Policy. In: The Wall Street Journal. Dow Jones and Company, Inc., New York (July 13, 2014).

Campo, E., Swerdlow, S.H., Harris, N.L., Pileri, S., Stein, H., Jaffe, E.S., 2011. The 2008 WHO classification of lymphoid neoplasms and beyond: evolving concepts and practical applications. Blood 117, 5019–5032 (As cited in Checkoway et al. 2015).

Checkoway, H., Dell, L.D., Boffetta, P., Gallager, A.E., Crawford, L., Lees, P.S.J., Mundt, K.A., 2015. Formaldehyde exposure and mortality risks from acute myeloid leukemia and other lymphohematopoietic malignancies in the US National Cancer Institute Cohort Study of Workers in Formaldehyde Industries. J. Occup. Environ. Med. 57 (7), 785–794.

Collins, F.S., Tabak, L.A., 2014. Policy: NIH plans to enhance reproducibility. Nature 505 (7485), 612–613.

ECHA, 2011. CLH Report Proposal for Harmonised Classification and Labelling. Based on Regulation (EC) No 1272/2008 (CLP Regulation), Annex VI, Part 2. Substance Name: Formaldehyde. European Chemicals Agency. September 28. Version No. 2.

Edrissi, B., Taghizadeh, K., Moeller, B.C., Kracko, D., Doyle-Eisele, M., Swenberg, J.A., Dedon, P.C., 2013. Dosimetry of N6-Formyllysine adducts following [13C2H2]-Formaldehyde exposures in rats. Chem. Res. Toxicol. 26 (10), 1421–1423.

Heron, M., Hoyert, D.L., Murphy, S.L., Xu, J., Kochanek, K.D., Tejada-Vera, B., 2006. Deaths: final data for 2006. United States Department of Health and Human Services, centers for disease control and prevention, national center for health statistics, national vital statistics system. Natl. Vital Stat. Rep. 57 (14), 135. http://www.cdc.gov/nchs/data/nvsr/nvsr57/nvsr57_14.pdf.

Hauptmann, M., Lubin, J.H., Stewart, P.A., Hayes, R.B., Blair, A., 2004. Mortality from solid cancers among workers in formaldehyde industries. Am. J. Epidemiol. 159 (12), 1117–1130.

IARC, 2012. Formaldehyde. Chemical Agents and Related Occupants. Volume 100F a Review of Human Carcinogens. IARC Monographs on the Evaluation of Carcinogenic Risks to Humans. International Agency for Research on Cancer. World Health Organization, Lyon, France.

Igbal, S.A., Wallach, J.D., Khoury, M.J., Schully, S.D., Ioannidis, J.P., 2016. Reproducible research practices and transparency across the Biomedical Literature. PLoS Biol. 14 (1), e1002333.

Jilka, R.L., 2016 Jul. The road to reproducibility in animal research. J. Bone Mineral Res. 31 (7), 1317–1319. http://dx.doi.org/10.1002/jbmr.2881.

Lu, K., Moeller, B., Doyle-Eisele, M., McDonald, J., Swenberg, J.A., 2011. Molecular dosimetry of N^2-hydroxymethyl-dG DNA adducts in rats exposed to formaldehyde. Chem. Res. Toxicol. 24 (2), 159–161.

Moeller, B.C., Lu, K., Doyle-Eisele, M., McDonald, J., Gigliotti, A., Swenberg, J.A., 2011. Determination of N^2-Hydroxymethyl-dG adducts in the nasal epithelium and bone marrow of nonhuman primates following $^{13}CD_2$-formaldehyde inhalation exposure. Chem. Res. Toxicol. 24, 162–164.

Muller-Hermelink, H.K., Montserrat, E., Catovsky, D., Harris, N.L., 2001. Chronic lymphocytic leukemia/small lymphocytic lymphoma. In: Jaffe, E., Harris, N., Stein, H., et al. (Eds.), Pathology and Genetics: Tumours of Haematopoietic and Lymphoid Tissues. IARC Press, Lyon, France, pp. 127–130 (As cited in Checkoway, et al. 2015).

NCHS 2009. (As cited in USEPA 2010 – citation was not reflected in the reference list).

NRC, 2011. Review of the Environmental Protection Agency's Draft IRIS Assessment of Formaldehyde. Committee to Review EPA's Draft IRIS Assessment of Formaldehyde Board on Environmental Studies and Toxicology Division on Earth and Life Studies. National Research Council of the National Academies. The National Academies Press, Washington, DC, p. 194.

NRC, 2014. Review of EPA's Integrated Risk Information System (IRIS) Process. Committee to Review the IRIS Process Board on Environmental Studies and Toxicology Division on Earth and Life Studies. National Research Council of the National Academies. The National Academies Press, Washington, DC, p. 204.

NTP, 2014. Formaldehyde (CAS No. 50-00-0). Report on Carcinogens, thirteenth ed. National Toxicology Program, Untied States Department of Health and Human Services, Public Health Service.

Stewart, P.A., Blair, A., Cubit, D.A., Bales, R.E., Kaplan, S.A., Ward, J., Gaffey, W., O'Berg, M.T., Walrath, J., 1986. Estimating historical exposures to formaldehyde in a retrospective mortality study. Appl. Ind. Hyg. 1 (1), 34–41.

USEPA, 2005. Guidelines for Carcinogen Risk Assessment. United States Environmental Protection Agency, Risk Assessment Forum, Washington, DC. EPA/630/P-03/001F.

USEPA, 2010. Toxicological Review of Formaldehyde-inhalation Assessment (CAS No. 50-00-0). In: Support of Summary Information on the Integrated Risk Information System. EPA/635/R-10/002A.

USEPA, 2012. Benchmark Dose Technical Guidance. Risk Assessment Forum, United States Environmental Protection Agency, Washington, DC. EPA/100/R-12/001.

Yu, R., Lai, Y., Hartwell, H.J., Moeller, B.C., Doyle-Eisele, M., Kracko, D., Bodnar, W.M., Starr, T.B., Swenberg, J.A., 2015. Formation, accumulation, and hydrolysis of endogenous and exogenous formaldehyde-induced DNA damage. Toxicol. Sci. 146 (1), 170–182.

Zhang, L., Steinmaus, C., Eastmond, D.A., Xin, X.K., Smith, M.T., 2009. Formaldehyde exposure and leukemia: a new meta-analysis and potential mechanisms. Mutat. Res. 681 (2–3), 150–168.

Zhang, L., Tang, X., Rothman, N., Vermeulen, R., Ji, Z., Shen, M., Qiu, C., Guo, W., Liu, S., Reiss, Beane Freeman L., Ge, Y., Hubbard, A.E., Hua, M., Blair, A., Moore, L.E., Kim, S., Xie, Y., Hayes, R.B., Azuma, M., Rappaport, S.M., Huang, H., Fraumeni Jr., J.F., Smith, M.T., Lan, Q., 2010. Occupational exposure to formaldehyde, hematotoxicity, and leukemia-specific chromosome changes in cultured myeloid progenitor cells. Cancer Epidemiol. Biomarkers, Prev. 19 (1), 80–88.